

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:52:45 ; Search time 170.72 seconds  
(without alignments)  
16.705 Million cell updates/sec

Title: US-09-331-631A-8\_COPY\_120\_161

Perfect score: 245  
Sequence: 1 SORFOECQCHQOEPRPEKOCVRECKRYQENPMRGER 42

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR66:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	245	100.0	509	2 S08059	alpha-globulin typ
2	231	94.3	588	1 FMCNAB	alpha-globulin B p
3	188	76.7	605	2 S06398	alpha-globulin typ
4	111.5	45.5	566	2 S22477	viciilin precursor
5	74	30.2	3828	2 T13857	trithorax protein
6	72	29.4	425	2 T18592	hypothetical prote
7	72	29.4	600	2 T18593	hypothetical prote
8	71	29.0	124	2 A37931	napiin - Swedish tu
9	71	29.0	127	2 S65447	allergen Sin a I -
10	70	28.6	637	2 S35221	globulin Bgl prec
11	70	28.6	905	2 T23229	hypothetical prote
12	68	27.8	930	2 T08588	hypothetical prote
13	66.5	27.1	106	2 S26336	hypothetical prote
14	66.5	27.1	152	2 PS0427	napiin ntb - rape
15	66.5	27.1	174	2 PS0425	napiin AHI precursor
16	66.5	27.1	255	2 A60637	napiin B43 precursor
17	66	26.9	816	2 S46268	meiozoite antigen
18	65	26.5	388	2 T31887	ataxin-1 - human
19	65	26.5	388	2 T31888	hypothetical prote
20	65	26.5	438	2 T31889	hypothetical prote
21	65	26.5	445	2 T31898	hypothetical prote
22	64.5	26.3	411	2 T29475	hypothetical prote
23	64.5	26.3	807	2 A32582	protein kinase YAK
24	64	26.1	186	2 A29802	napiin precursor (g
25	64	26.1	242	2 T29699	hypothetical prote
26	64	26.1	305	2 S08312	gamma-hordein 1 pr
27	64	26.1	838	2 T20125	hypothetical prote
28	64	26.1	905	1 RGRYS5	regulatory protein
29	64	26.1	1898	1 A45973	trichomyalin - hum

30	63.5	25.9	110	2 S20350	napiin nla - rape
31	63.5	25.9	145	2 PC1247	Sin a I allergen 2
32	63.5	25.9	145	2 PC1246	Sin a I allergen 1
33	63.5	25.9	145	2 S65481	allergen Sin a I
34	63.5	25.9	145	2 S65482	allergen Sin a I
35	63.5	25.9	145	2 S65480	allergen Sin a I
36	63.5	25.9	145	2 S65479	allergen Sin a I
37	63.5	25.9	155	2 PS0426	napiin B69 precursor
38	63.5	25.9	178	1 NMRP2	napiin 2 precursor
39	63.5	25.9	178	2 S07828	napiin B - rape
40	63.5	25.9	178	2 S15382	napiin napB - rape
41	63.5	25.9	178	2 A25997	napiin precursor (n
42	63.5	25.9	178	2 S25130	2S storage protein
43	63.5	25.9	180	2 S10018	napiin (clone BqNA
44	63.5	25.9	180	2 S52025	napiin (clones BMM
45	63.5	25.9	521	2 JC5220	protein-tyrosine-p

## ALIGNMENTS

RESULT 1  
S08059  
alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)  
N:Alternate names: seed storage protein  
C:Species: Gossypium hirsutum (upland cotton)  
C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 30-Sep-1993  
C:Accession: S08059  
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A>Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.  
A:Reference number: S06398  
A:Accession: S08059  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-509 <CHL>  
C:Superfamily: glycine

Query Match 100.0%; Score 245; DB 2; Length 509;  
Best local Similarity 100.0%; Pred. No. 1.1e-19;  
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SORFOECQCHQOEPRPEKOCVRECKRYQENPMRGER 42  
DB 39 SORFOECQCHQOEPRPEKOCVRECKRYQENPMRGER 80  
|||||  
FMCNAB  
alpha-globulin B precursor (clone C72) - upland cotton  
N:Alternate names: seed storage protein; viciilin precursor  
C:Species: Gossypium hirsutum (upland cotton)  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999  
C:Accession: A30838; S06911  
R:Chlan, C.A.; Pyle, J.B.; Legocki, A.B.; Dure III, L.  
Plant Mol. Biol. 7, 475-489, 1986  
A>Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII  
A:Reference number: A30838  
A:Accession: A30838  
A:Molecule type: mRNA  
A:Residues: 1-588 <CHL>  
A:Cross-references: GB:M16891; NID:G167374; PIDN:AAA3071.1; PID:G167375  
A:Experimental source: var. Coker 201  
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A>Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX.  
A:Reference number: S06398  
A:Accession: S06911  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-81 <CH2>  
C:Comment: This is a seed storage protein.

C:Superfamily: glycinin  
C:Keywords: glycoprotein; seed; storage protein  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-588/Product: alpha-globulin storage protein #status predicted <MAT>  
F:417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 94.3%; Score 231; DB 1; Length 588;  
Best Local Similarity 92.9%; Pred. No. 4.2e-18;  
Matches 39; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SQRFQECQCHQOEORPEKKQOCVRECKRYOENPMRGER 42  
Db 120 SQRFQECQCHQOEORPEKKQOCVRECKRYOENPMRER 161

RESULT 3  
S06398  
alpha-globulin type A precursor - upland cotton  
N:Alternate names: seed storage protein  
C:Species: Gossypium hirsutum (upland cotton)  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 30-Sep-1993  
C:Accession: S06398  
R:Chlan, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.  
Plant Mol. Biol. 9, 533-546, 1987  
A:Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se  
A:Reference number: S06398  
A:Accession: S06398  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
C:Superfamily: glycinin  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-605/Product: alpha-globulin type A #status predicted <MAT>

Query Match 76.7%; Score 188; DB 2; Length 605;  
Best Local Similarity 75.0%; Pred. No. 2.2e-13;  
Matches 30; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 3 ROFQECQCHQOEORPEKKQOCVRECKRYOENPMRGER 42  
Db 120 KQFQECQCHQOEORPEKKQOCVRECKRYOENPMKGER 159

RESULT 4  
S22477  
vicilin precursor - cacao  
C:Species: Theobroma cacao (cacao)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Mar-1999  
C:Accession: S22477; S22478; S18105; S22050  
R:McHenry, L.; Fritz, P.J.  
Plant Mol. Biol. 18, 1173-1176, 1992  
A:Title: Comparison of the structure and nucleotide sequence of vicilin genes of cocoa  
A:Reference number: S22477; MUID:92288309  
A:Accession: S22477  
A:Molecule type: DNA  
A:Residues: 1-566 <MCCH>  
A:Cross-References: EMBL:X62625  
A:Accession: S22478  
A:Molecule type: mRNA  
A:Residues: 1-452 <MC2>  
A:Cross-References: EMBL:X62626  
C:Genetics:  
A:Introns: 211/1; 269/3; 296/3; 391/3; 502/1  
C:Superfamily: glycinin  
C:Keywords: seed; storage protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-566/Product: vicilin #status predicted <MAT>

Query Match 45.5%; Score 111.5; DB 2; Length 566;  
Best Local Similarity 50.0%; Pred. No. 3e-05;

Matches 20; Conservative 9; Mismatches 10; Indels 1;  
QY 2 QRFQECQCHQOEORPEKKQOCVRECKRYOENPMRGE 41  
Db 82 QRFQECQCHQOEORPEKKQOCVRECKRYOENPMRGE 120

RESULT 5  
T13857  
trithorax protein - fruit fly (Drosophila virilis)  
C:Species: Drosophila virilis  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
C:Accession: T13857  
R:Mazo, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: T13857  
A:Accession: T13857  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3828 <MA2>  
A:Cross-References: EMBL:Z50038; NID:g899253; PID:g899254; PIDN:CAA90349.1  
C:Genetics:  
A:Introns: 337/3; 529/1; 721/1; 791/1; 3668/2; 3713/1; 3771/3  
C:Superfamily: Drosophila trithorax protein  
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 30.2%; Score 74; DB 2; Length 3828;  
Best Local Similarity 38.2%; Pred. No. 3.1;  
Matches 13; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRFQECQCHQOEORPEKKQOCVRECKRYOE 35  
Db 2995 QRFQECQCHQOEORPEKKQOCVRECKRYOE 3028

RESULT 6  
T18592  
hypothetical protein AC3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T18592  
R:McMurray, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: T18592  
A:Accession: T18592  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-425 <MIL2>  
A:Cross-References: EMBL:Z71177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3  
A:Experimental source: clone AC3  
C:Genetics:  
A:Gene: CESP:AC3.3  
A:Map position: 5  
A:Introns: 18/3  
C:Superfamily: gliadin

Query Match 29.4%; Score 72; DB 2; Length 425;  
Best Local Similarity 46.4%; Pred. No. 0.83;  
Matches 13; Conservative 7; Mismatches 4; Indels 4; Gaps 1;

QY 6 QRFQECQCHQOEORPEKKQOCVRECKRYOE 29  
Db 157 QRFQECQCHQOEORPEKKQOCVRECKRYOE 184

RESULT 7  
T18593  
hypothetical protein AC3.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18593



F:290-324/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:329-364/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:431-465/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:513-550/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
 F:579-614/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
 F:767-801/Domain: LDL receptor ligand-binding repeat homology <LDL7>

Query Match 28.6%; Score 70; DB 2; Length 905;  
 Best Local Similarity 39.4%; Pred. No. 2.6;  
 Matches 13; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

OY 4 OROFOECQOHQOORPEKKQOCVRECKRYOEN 36  
 DB 645 OAOQAQOHQOORQOORQOORQOORQOORQOOD 677

## RESULT 12

hypothetical protein L23H3.30 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 22-Oct-1999  
 C:Accession: T08588

R:Bevan, M.; Pohl, T.; Weizemagter, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schuel  
 submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16098

A:Accession: T08588

A:Molecule type: DNA

A:Residues: 1-930 <BEV>

A:Cross-references: EMBL:AL050398; GSPDB:GN00062; ATSP:L23H3.30

A:Experimental source: cultivar Columbia; BAC clone L23H3

C:Genetics:

A:Gene: ATSP:L23H3.30

A:Map position: 4

A:introns: 11/2; 51/1; 87/3; 249/3; 278/2; 304/3; 330/1; 346/3; 449/3; 523/3; 605/3; 645

Query Match 27.8%; Score 68; DB 2; Length 930;  
 Best Local Similarity 36.8%; Pred. No. 4.4;  
 Matches 14; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

OY 2 OROFOECQOHCH--OOORPEKKQOCVRECKRYOENP 37  
 DB 130 QQQQQQQQQHHHHQQQQQQQQQQQQQQQQQQHQNQP 167

## RESULT 13

S26636

napin n1b - rape

C:Species: Brassica napus (rape)

C:Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 04-Oct-1996

C:Accession: S26636

R:Monalve, R.I.; Lopez-Otin, C.; Villalba, M.; Rodriguez, R.

FEBS Lett. 295, 207-210, 1991

A:Title: A new distinct group of 2 S albumins from rapeseed. Amino acid sequence of two

A:Reference number: S20350; MUID:92111741

A:Accession: S26636

A:Molecule type: Protein

A:Residues: 1-106 <MON>

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: pyroglyutamic acid; seed; storage protein

F:1-31/Product: napin small chain #status experimental <SMA>

F:32-106/Product: napin large chain #status experimental <LAR>

F:32/Modified site: pyroliidone carboxylic acid (Gln) (in mature form) #status experime

Query Match 27.1%; Score 66.5; DB 2; Length 106;  
 Best Local Similarity 35.3%; Pred. No. 1;  
 Matches 18; Conservative 6; Mismatches 8; Indels 19; Gaps 3;

OY 2 OROFOE-----COQHCHQO-----EOPEKKQOCVRECKRYOEN 36  
 DB 6 QREFOQOHLRACQOWIRQOLAGSPQSGPQGPWLREOC---CNELYOED 53

RESULT 14  
 PS0427  
 napin AH1 precursor - radish (fragment)

C:Species: Raphanus sativus (radish)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999

C:Accession: PS0427

R:Raynal, M.; Depigny, D.; Grellet, F.; Delseny, M.

Gene 99, 77-86, 1991

A:Title: Characterization and evolution of napin-encoding genes in radish and related

A:Reference number: PS0425; MUID:91216448

A:Accession: PS0427

A:Molecule type: mRNA

A:Residues: 1-152 <RAY>

A:Cross-references: GB:M63842; GB:M63629; NID:g169695; PIDN:AAA63471.1; PID:g169696

A:Experimental source: seed

C:Superfamily: wheat alpha-amylase inhibitor

F:1-6/Domain: signal sequence and propeptide (fragment) #status predicted <PRO>

F:7-41/Product: napin small chain #status predicted <NSC>

F:61-146/Product: napin large chain #status predicted <MLC>

Query Match 27.1%; Score 66.5; DB 2; Length 152;  
 Best Local Similarity 27.1%; Pred. No. 1.4;  
 Matches 19; Conservative 6; Mismatches 8; Indels 37; Gaps 3;

OY 2 OROFOE-----COQHCHQO-----EOPEKKQOCVR 27  
 DB 17 RREFQOQHLRACQOWLHQARQSGSPWTLDDERFEDDMENPOGPOQRPPLLDQC--- 74

OY 28 ECKREKYOENP 37  
 DB 75 -CNELHQOEEP 83

## RESULT 15

PS0425

napin BA3 precursor - radish (fragment)

C:Species: Raphanus sativus (radish)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 16-Jul-1999

C:Accession: PS0425

R:Raynal, M.; Depigny, D.; Grellet, F.; Delseny, M.

Gene 99, 77-86, 1991

A:Title: Characterization and evolution of napin-encoding genes in radish and related

A:Reference number: PS0425; MUID:91216448

A:Accession: PS0425

A:Molecule type: mRNA

A:Residues: 1-174 <RAY>

A:Cross-references: GB:M63843; GB:M63630; NID:g169697; PIDN:AAA63472.1; PID:g169698

A:Experimental source: seed

C:Superfamily: wheat alpha-amylase inhibitor

C:Keywords: storage protein

F:1-28/Domain: signal sequence and propeptide (fragment) #status predicted <PRO>

F:29-63/Product: napin small chain #status predicted <NSC>

F:83-168/Product: napin large chain #status predicted <MLC>

Query Match 27.1%; Score 66.5; DB 2; Length 174;  
 Best Local Similarity 27.1%; Pred. No. 1.6;  
 Matches 19; Conservative 6; Mismatches 8; Indels 37; Gaps 3;

OY 2 OROFOE-----COQHCHQO-----EOPEKKQOCVR 27  
 DB 39 RREFQOQHLRACQOWLHQARQSGSPWTLDDERFEDDMENPOGPOQRPPLLDQC--- 96

OY 28 ECKREKYOENP 37  
 DB 97 -CNELHQOEEP 105

Search completed: March 1, 2001, 15:52:46

Job time: 570 sec

---

